PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

rstb.royalsocietypublishing.org

Opinion piece

Cite this article: Nolan F, Jeon H-S. 2014 Speech rhythm: a metaphor? *Phil. Trans. R. Soc. B* **369**: 20130396.
http://dx.doi.org/10.1098/rstb.2013.0396

One contribution of 14 to a Theme Issue 'Communicative rhythms in brain and behaviour'.

Subject Areas:
behaviour

Keywords:
speech rhythm, speech timing, rhythm metrics, prosodic prominence, tune–text association

Author for correspondence:
Francis Nolan
e-mail: fjn1@cam.ac.uk

†Present address: International Institute of Korean Studies, School of Language, Literature and International Studies, University of Central Lancashire, Preston, PR1 7BH, UK.

Royal Society Publishing

Speech rhythm: a metaphor?

Francis Nolan and Hae-Sung Jeon†

Phonetics Laboratory, Department of Theoretical and Applied Linguistics, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DA, UK

Is speech rhythmic? In the absence of evidence for a traditional view that languages strive to coordinate either syllables or stress-feet with regular time intervals, we consider the alternative that languages exhibit *contrastive rhythm* subsisting merely in the alternation of stronger and weaker elements. This is initially plausible, particularly for languages with a steep 'prominence gradient', i.e. a large disparity between stronger and weaker elements; but we point out that alternation is poorly achieved even by a 'stress-timed' language such as English, and, historically, languages have conspicuously failed to adopt simple phonological remedies that would ensure alternation. Languages seem more concerned to allow 'syntagmatic contrast' between successive units and to use durational effects to support linguistic functions than to facilitate rhythm. Furthermore, some languages (e.g. Tamil, Korean) lack the lexical prominence which would most straightforwardly underpin prominence of alternation. We conclude that speech is not incontestibly rhythmic, and may even be antirhythmic. However, its linguistic structure and patterning allow the metaphorical extension of rhythm in varying degrees and in different ways depending on the language, and it is this analogical process which allows speech to be matched to external rhythms.

1. Introduction

In this paper, we will put forward a somewhat iconoclastic view of speech rhythm. As a preliminary, however, we consider how far the terms *rhythm* and *timing* are distinct, how far they overlap and how uncontroversially rhythm applies to speech. Of the two terms, 'timing' is by far the simpler. Admittedly, if you are a theoretical physicist, you can make the notion of time difficult; but in the experience of ordinary mortals life is based around clocks of various kinds—whether in the natural world, such as the diurnal cycle, or man-made—which tick away the span between birth and death in larger or smaller units. Humans are now masters of time, in the sense that we can measure the duration of things to a delicacy unimaginable historically; but we remain slaves of time, in that we can neither travel through it nor stop it.

As far as speech is concerned, there can be no controversy over the fact that time is of the essence. Any attempt to synthesize speech without taking the subtle timing relations that permeate speech into account will result in an output which is unnatural and potentially unintelligible. From the duration of acoustic events such as plosive aspiration which contribute to the phonetic exponence of phonological segments, to the lengthening often cuing the boundaries of prosodic phrases, time is crucial. None of this, of course, is straightforward; modelling the orderly use of time in the different layers of structure that make up speech is an immensely challenging task. But on the whole, the task is at least well defined: measure durations, hypothesize models (which by the nature of speech are likely to be hierarchical) and test them against data.

When we come to rhythm in speech, the matter is much more open to controversy. We will entertain the possibility that there is no such thing as speech rhythm, and even that speech is inherently antirhythmic. Nonetheless, there is a rich tradition of trying to describe speech rhythm—which implies that this iconoclastic stance has not been widely shared. To debate the matter, we need to consider what rhythm is. There seem to be two broad ways to approach the definition of rhythm, one which emphasizes regularity in time, and the other which emphasizes structural relations. The first type of rhythm can be termed

coordinative rhythm [1], also often known as *periodic* rhythm [2]; it implies both repetition of a pattern and regularity of the interval taken by each repetition. This conceptualization has been called the *temporal* view of rhythm [3]. Examples of such rhythm abound in the physical world: a train travelling at steady speed over jointed track, producing the classic 'clickety-clack, clickety-clack'; the sound of sawing, once the stroke has been established; and the successive cycles of the beating heart. Western music of a type which allows foot-tapping or hand-clapping fulfils this definition, with the organization of notes into bars which are isochronous—give or take some deviation from regularity on the part of a skilled performer for artistic effect. All of these are characterized by regularity in time, or at least an approximation to it. There is co-extension of the recurring event with a specific interval measurable by an external 'clock', or co-occurrence of a definable point in the event with equally spaced points in time as determined by that external clock. In the case of speech, coordinative rhythm would arise from the organization of sounds into groups marked by phonetic cues and synchronized in time with the objective regularity. Strictly, coordinative rhythm therefore entails what has often been called in the speech literature *isochrony*, meaning that a given repeated element or structural grouping of elements (e.g. syllable or foot) should always occupy the same time span. In the case of a group whose elements may vary in size or number, compensatory adjustments in durations would be needed to make those elements 'fit'.

Coordinative rhythm is classically differentiated [1,2] from *contrastive* rhythm, a view which sees rhythm as consisting in the alternation of stronger and weaker elements (for a finer grained set of distinctions see [4]). Those stronger and weaker elements are constrained therefore by sequencing, and their strength or weakness may involve relative durations, but neither they nor the groups into which they are organized need be synchronized to an objective external clock. In the case of a language such as English, it is natural to map this 'non-temporal' [3] definition of rhythm onto the alternation of stressed and unstressed syllables. This is in accord with Patel [5] who suggests that progress will be made 'if one thinks of [speech] rhythm as systematic timing, accentuation, and grouping patterns in a language *that may have nothing to do with isochrony*'. On the contrastive view, the clickety-clack of the train, successful sawing and the beating heart are likewise rhythmic, consisting as they do of alternating distinct events. But presumably when the sawyer hits a knot in the wood and struggles to complete each stroke so that the even intervals in the cycle break down, the event remains rhythmic under the contrastive view, but not the coordinative (periodic or temporal) view. When the heart behaves in a correspondingly erratic fashion, we get what is tellingly termed 'cardiac arhythmia'. The patient with palpitations may or may not be comforted to know that as long as systoles and diastoles are broadly alternating, the heart's behaviour still fulfils the contrastive definition of rhythm.

In what follows we will comment on selected aspects of the study of speech rhythm within (and beyond) the context of dichotomy between coordinative and contrastive rhythm. First, we will deal with existing evidence against isochrony in speech and review briefly the history and application of rhythm 'metrics' to speech, including the question of language discrimination by babies as well as adults. We then discuss the notion of the 'prominence gradient' inherent in one metric, the pairwise variability index (PVI), and suggest with the benefit of hindsight that it constituted an explicit recognition of the role of *contrastive* rhythm in speech. Permeating the discussion will be a second theme, namely the extent to which it is appropriate to give 'timing' exclusive status in the context of rhythm. The less rhythm involves synchronization with an external clock, and the more the phenomena in speech to which the term rhythm is applied involve patterns of prominence, the more we need to concern ourselves with other dimensions which might contribute to prominence. We will therefore refer briefly to an example of work which looks explicitly at the role of pitch in rhythm. The widespread hunt for rhythm in speech does not, of course, prove that speech is rhythmic, and we go on to put evidence to the contrary (see [4] whose authors also question this assumption). We then argue that the nature of language promotes arhythmicity in speech; and we suggest an alternative perspective on the relation of speech to rhythm. We question the natural assumption that the possibility of associating speech with an external rhythm (such as that of music), and the apparently tight constraints on that process, necessarily support the premise that speech is normally rhythmic; or that such constraints apply equally in all languages. Instead, we claim that speech is 'rhythmic' only metaphorically, and that the metaphor works better in some languages than others.

2. Coordinative rhythm and responses to the lack of isochrony

If the coordinative rhythm hypothesis involves the alignment of speech with an external regularity, it makes sense to ask what elements or structures in speech are so to be aligned. The answer, according to the seminal views of Pike [6] and Abercrombie [7], is that it depends on the language. In particular, it could be the syllable, as in French, or the stress-foot, as in English. These alternatives became widely known as *syllable-timing* and *stress-timing*, and were intuitively appropriate to capture a perceived prosodic difference between sets of languages which seemed to correspond with one possibility or the other. Syllables would all be the same length (isochronous) in the former type of language, and feet would be isochronous in the latter. That conceptualization is widely known and influential, as is by now the stubborn refusal of the data in a variety of languages to offer up straightforward confirmation of isochrony (e.g. [8]; though see [9] for a nuanced account of the history).

Historically, there were a number of responses to the disappointing outcome to the quest for isochrony as validation of the *syllable-timed*~*stress-timed* dichotomies which had so crisply focused thinking on speech rhythm; and by no means did the failure of the quest bring about a decline in interest in speech rhythm. Disobliging acoustic data could be circumvented by attributing isochrony to perception rather than the speech signal—in other words, relegating isochrony to the mental rather than acoustic sphere; Lehiste [10,11], for instance, cites a variety of evidence to support her claim that it is 'quite likely that the listener imposes a rhythmic structure on sequences of interstress intervals in spite of the fact that their durational differences are well above threshold' [10]. This, incidentally, reveals that Lehiste is thinking in terms of *coordinative* rhythm. Alternatively, rhythmic types were seen (e.g. [12,13]) as a function of the collaboration of the phonological, phonetic

and phonotactic properties of different languages, particularly syllable structure, rather than the result of the timing of intervals. To the extent that such views depended on (for instance) the homogeneity or diversity of successive syllables, they probably corresponded more to the *contrastive* component of rhythm, where it is the sequencing, and in particular alternation, of distinct units or structures, less or more homogeneous, which determines rhythmic type.

Despite the lack of quantitative evidence from the acoustic signal for a rhythmic dichotomy, and also a lack of empirical support for its reality in the perception of non-linguists and among speakers of different languages [4,14], remarks continued to be made about how this or that language or dialect was more *syllable-timed* or more *stress-timed* than another. From the 1990s, a number of 'rhythm metrics' or indices were developed which attempted to quantify the timing properties of languages that might be giving rise to such global impressions. Notable among these were the measures $\%V$ (the proportion of vocalic intervals in an utterance), ΔV (the standard deviation of duration of vocalic intervals) and ΔC (the standard deviation of duration of consonantal intervals) of Ramus [15,16], and Dellwo's development of the latter two as VarcoV and VarcoC [17,18]. These capture global, statistical trends of speech samples. In addition, comments on how Singapore English was nearer than British English to having syllable-timing or a 'staccato' rhythm [19,20] gave rise to another of the widely adopted rhythm metrics, the PVI of Low [21,22]. This too by-passed isochrony, and focused on the degree of variability between successive acoustic segments or phonological units, typically vocalic intervals or syllables. The difference in value v for a property p (for instance duration) between the members of each successive pair of phonological units throughout the speech sample is evaluated. A large average pairwise difference will reflect lack of regularity from unit to unit, and a small average will reflect regularity. In its simplest form, this 'raw' PVI leaves a problem because as values v for the property p are scaled up (for instance, as durations become proportionately greater under rallentando, or when a slower speaker is measured) the pairwise differences will become larger in a way which will falsely imply increasing irregularity. Low [21], therefore, used a *normalized PVI* as shown in equation (2.1) below, where each difference is calculated as a proportion of the mean value for p within the pair, the resultant fractional average PVI values being summed over the utterance, and multiplied by 100 merely to yield a whole number

$$\mathrm{PVI} = 100 \times \left[\sum_{k=2}^{n} \left| \frac{v_k - v_{k-1}}{(v_k + v_{k-1})/2} \right| / (n-1) \right]. \tag{2.1}$$

Lower normalized vowel PVI values (nPVI-V) were indeed found in Singapore English than British English [21,22]. This lent quantitative support for the implication of the 'syllable timed' label that syllables (represented by measurements of their vowels) show less variability in Singapore English.

The PVI was extended to cross-language comparison by Grabe & Low [23]. They took one speaker from each of 18 languages, and for duration calculated both the normalized vocalic PVI and the raw consonantal PVI (i.e. the variability in successive intervocalic intervals), plotting the disposition of the languages in a plane defined by those two parameters. Their data 'support a weak categorical distinction between stress-timing and syllable-timing … [but] … there is considerable overlap between the stress-timed and the syllable-timed group and hitherto unclassified languages' [23]. Numerous other studies have applied the PVI to different languages (e.g. [24–27]) and varieties including non-native accents (e.g. [28,29]).

A large number of criticisms have been levelled at rhythm metrics, targeting the technical details of the computations, their susceptibility to instability in the face of variation in factors such as speech rate, within-speaker variation, and measurement uncertainty, and their failure to capture the true nature of speech rhythm. To sample a few of these, Deterding [30], applying the PVI to Singapore English at the level of the syllable (as opposed to the more common vocalic intervals), noted that the purely pairwise normalization of the formula above would be sensitive to errors in the exact location of syllable boundaries and suggested normalization with respect to the mean syllable duration of the whole utterance. This remedy, however, would not answer the different objection of Gibbon [31] that 'alternating sequences may receive the same PVI as exponentially increasing or decreasing series PVI(2,4,2,4,2,4) = PVI(2,4,8,16,32,64)'. While this observation is correct, we are not aware of linguistic behaviour of this type actually occurring.

Knight [32] checked seven rhythm metrics for stability over time. An English passage of around 140 syllables was read on three successive days by four speakers. There were no significant differences over time in any of the metrics, with ΔV (the standard deviation of vowel durations) and $\%V$ (the proportion of vowel-time in the utterance) showing the best correlations over the 3 days (at 0.87 and 0.79), but nPVI-V also achieved a value (0.62) which could be considered, for the statistic used, as indicating 'substantial' stability. Less promisingly, dividing the passage into four subsections showed significant differences for all but one (VarcoC). However, this instability probably indicates, as acknowledged in Knight [32], that the subsections were too short to neutralize for phonological content and to allow the values of the metrics to 'settle'.

Gibbon [31] claims that 'The [PVI] model has an empirical problem: it assumes strictly binary rhythm. Hence, alternations as in *Little John met Robin Hood and so the merrie men were born* are adequately modelled, but not the unary rhythm (syllable timing) of *This one big fat bear swam fast near Jane's boat* or ternary dactylic and anapaestic rhythms (or those with even higher cardinality) like *Jonathan Appleby wandered around with a tune on his lips and saw Jennifer Middleton playing a xylophone down on the market-place*.' In fact, the originators of the PVI were perfectly aware that English rhythm is not consistently binary; Nolan [33] explicitly anticipates the point, as reprised in §6 below, and Lin & Wang [34] deliberately incorporate constructed sentences with 'unary' rhythm in the expectation that these would not differ in PVI between British and Singaporean English. The PVI metric does not 'assume binary rhythm', it assumes a sufficient predominance of strong-weak alternation in natural usage that the cumulative effect will be to raise the PVI value in a language impressionistically described as stress-timed. For instance, even in Gibbon's carefully constructed dactylic example (*Jonathan Appleby…*) with its 35 pairwise intervals, two-thirds (24) pair strong and weak syllables (assuming secondary stress on '-place').

In a comprehensive cross-linguistic study using several metrics, Arvaniti [27] shows, as would be hoped, a significant main effect of language, but only patchy correlations between

different rhythm metrics, and considerable instability in the values depending on factors such as speaking style (read sentences, read story or spontaneous speech), between-speaker variation, and (unsurprisingly, as metrics rely on the distributional statistics of a sample settling over a sufficient large and random sample) the structure of sentences deliberately designed to be maximally 'stress-timed' or maximally 'syllable-timed' within the constraints of each language. Broadly, the conclusion is that rhythm metrics are poorly capturing rhythm classes for which in any case there is no independent evidence other than the impressions of some linguists. Arvaniti's view [14] is that 'it appears advantageous to adopt a conception of rhythm that goes beyond timing and rhythmic types but rests instead on grouping and patterns of prominence. In this respect, connecting phonetic research to models of rhythm that are widely accepted in phonology ... and closer to the psychological understanding of rhythm may also be beneficial', though no specific proposal is made for the replacement of rhythm metrics with a concise and less imperfect alternative in the domain of cross-language comparison of global rhythm profiles. We will return to the implication that rhythm metrics have only been about timing in §4.

3. Perceptual discrimination of languages through rhythm

While phonological models of rhythm may be relevant to native speakers' perceptions of their languages, their knowledge of phonological and prosodic structure is acquired, not innate. Early on, rhythm metrics were applied not only to adult but also pre-linguistic child listeners' ability for rhythmic categorization (see [1,35] for reviews) and showed some predictive power with respect to listeners' discrimination performance (e.g. [15,36]). A series of experiments with babies between birth and five months old as well as adults demonstrated that they could discriminate rhythmically different languages even with low-pass filtered speech lacking phonemic and phonotactic information or resynthesized speech which does not include any segmental and intonational information. For instance, French adults discriminated so-called 'stress-timed' English from mora-timed Japanese with only durational cues [37]; French babies discriminated English from Japanese, syllable-timed Spanish, or Italian, but not from stress-timed Dutch without segmental information [38], and discriminated Dutch from Japanese with only durational information [39]. In particular, results with babies, who do not have the knowledge of their native language phonology and lexicon, show that they attend to prosodic similarities or differences in speech.

It is sometimes argued that such sensitivity to rhythm and putative *rhythm classes* on the part of babies underpins the acquisition of speech segmentation strategies (for example, English-speaking babies learn to group consecutive syllables into prominence-headed word-like units). That is, the findings have been interpreted to support the presence of discrete rhythm 'categories' or the possible existence of more rhythm categories yet to be found (e.g. [35]). However, 'the psychological reality of rhythm classes' [16] should not be given as the only possible interpretation. The overall durational properties over an utterance are determined by a range of segmental and supra-segmental factors (cf. [40]), such as intrinsic segmental duration, syllable structure, lengthening or shortening at the edge of word or prosodic unit, durational adjustment related to the distribution or relation of stress or accent, and speech rate, not solely based on the language's 'rhythm unit'. Languages and dialects differ in the presence and/or the extent of each durational process, i.e. there are multiple sources of the cross-linguistic or -dialectal prosodic differences.

Furthermore, the claim that listeners do not discriminate languages in the same rhythm category and do discriminate languages in different rhythmic categories begs the question of whose perception we are dealing with. The syllable-timed and stress-timed dichotomies were created by English speakers (e.g. [41] for a critique), and recent evidence suggests that the perceptual integration of cues to rhythm, centrally duration and pitch, is not universal ([42]; see §5 below for a summary), implying that such categorization experiments may not always show cross-linguistically valid results. In addition to their native language, listeners' discrimination performance seems to be affected by the nature of the materials presented and their maturational stage. For instance, English-speaking adult listeners discriminate English dialects in the same rhythm category with durational cues alone [1]. In Christophe & Morton [43], which used a habituation/dishabituation paradigm, 2- to 3-month-old English babies did not discriminate foreign language sentences supposedly in different rhythm categories, specifically mora-timed Japanese and syllable-timed French; and only weak discrimination performance was observed for the mora-timed Japanese and stress-timed Dutch pair. The findings on the rhythmic category discrimination can be explained without assuming the existence of discrete rhythm classes, as some languages or dialects can be prosodically more similar to each other than the others and listeners would attend to salient differences. However, to the extent that pre-linguistic listeners are able to discriminate speech samples differing only in timing, support is given to the relevance of rhythm metrics, as such listeners are presumably not yet in a position to apply higher level judgements about timing effects determined by the hierarchical prosodic structure of different languages. In this connection, the innovative approach of Tilsen & Arvaniti [44], which involves the automatic extraction from the speech signal of metrics from the amplitude envelope, including periodicities of higher (syllable) and lower (stress) frequencies, and their relative strengths, may provide an alternative model for the perception of babies who as yet have no knowledge of the phonological structure of languages.

Studies explicitly comparing the performance of different metrics in terms of their ability to discriminate languages traditionally thought of as rhythmically distinct (e.g. [35,45–48]) have generally shown that the metrics all had at least some areas of success explicating global intuitions about rhythmic differences. However, the linguistic sources of durational variations are very difficult, or even impossible to infer from global metrics, particularly when the materials are not carefully designed (cf. [27]). In sum, the outcome of such studies is that rhythm metrics do capture differences between languages, albeit imperfectly, and to some extent these differences correspond to syllable-timing versus stress-timing; but the results show no support for discrete categories as opposed to one or more continua along which languages can range themselves, and nor do they demonstrate that the perceived differences are the result of either a rhythmic intent on the part of the speaker or a cyclicity underlying the process of speech production.

4. The prominence gradient

What is often neglected in the literature on rhythm metrics, including the critiques cited above, is that the PVI, alone among contemporaneous metrics, was not specific to timing/duration. Given the importance of, for instance, vowel reduction in English as a factor contributing to its perceived rhythm, the original conception of the PVI [21] was that it could be calculated on any parameter relevant to prominence. In Low [21], RMS amplitude as well as duration PVIs were calculated, together with a measure of vowel spectral dispersion [22], all showing lower variability in Singapore English than British English. It is true that the PVI set a hare running, in pursuit of which this multi-dimensionality was lost sight of in favour of a focus on timing; but the PVI is, at core, both neutral between dimensions and multi-dimensional. More recently, Ferragne [49] incorporated syllable intensity in an application of the PVI in an automatic method for attributing British English samples to a set of accent (dialect) categories. The combination gave better discrimination of the accents than duration alone, and recognizes that whatever characterizes rhythmic differentiation between accents it does not depend just on timing. Cumming [50] explicitly weighs duration and F0, in a combined PVI, on the basis of the respective sensitivity of Swiss-French and Swiss-German listeners to these parameters when judging rhythmicality.

The rationale for conceiving the PVI as equally applicable to multiple dimensions has subsequently been formalized in the notion of the *prominence gradient* [26]. In all languages there will be at least some variation between the prominence of successive syllables, prominence being (for these purposes) the percept arising from the cumulative effect of high values in duration, intensity, pitch obtrusion (often, but not inevitably, upwards), vowel spectral dispersion and spectral balance. Different languages will employ different weightings of these (see for instance [50–52]). In some languages, there will be relatively sharp steps down or up in prominence between successive syllables—giving on average a steep mean pairwise prominence gradient; in others the gradient will be much shallower. The concept of a dichotomy, or continuum, between stress-timing and syllable-timing is thus replaced by a continuum based on global properties of prominence alternation within a language. In attempting to capture trends in the prominence gradient, the PVI has a clear affinity with the contrastive view of speech rhythm and, therefore, a theoretical foundation which we believe is more explicit than in the case of other indices. However, neither this affinity, nor the fact that a PVI value can always be generated, resolves the question of whether speech is rhythmic by design, as will be discussed in §7.

5. The interaction of timing and pitch

A prerequisite for profitably incorporating multiple dimensions into the study of rhythm is a better understanding of how those dimensions are integrated and how they interact. Barry *et al.* [53], for instance, show that for German listeners, differences in F0 can achieve a percept of rhythmicality more effectively than durational alternation. Cumming [42], in similar vein, presents a study exploring interactions between perceived duration and F0 in judging the naturalness of speech rhythm in short samples of connected speech. Sentences which were designed to be as comparable as possible structurally in Swiss-German and Swiss-French had duration and F0 manipulated by resynthesis. Specifically, in sentences such as

De Leerer | wiirt d**Schüe**ler | erchäne |
The teacher *will the pupils* *recognize*
L'enseignante | a con**nu** | les élèves |
The teacher *knew* *the pupils*

the syllable in bold was increased or decreased from the original by 35% in duration and three semitones in pitch, yielding nine versions including those replicating original values in each dimension. Subjects were presented with a 3×3 randomized array of the stimuli for each sentence, and had to listen to all until they could make a decision as to which was 'rhythmically most natural'.

The experiment revealed that in judging rhythmic appropriateness in manipulated speech, Swiss-German speakers are intolerant of durational manipulation of a crucial syllable, while accepting some increase or decrease in its pitch range; whereas speakers of Swiss-French are tolerant of shorter duration and smaller pitch range, but less so of longer duration and/or increased pitch range. Speakers of standard French showed similar results. The most striking finding is in the reaction to durational manipulations; the speakers of Swiss-German were much more sensitive than those of French to 'wrong' durations of the crucial syllable.

There are of course difficulties with the precise interpretation of this, given that it is impossible to construct utterances in two languages which are perfectly matched for structure. Nonetheless, the experiment shows that where the task is to judge rhythm, subjects are sensitive to pitch as well as timing, and, less predictably, that the integration of these dimensions may well be language-specific. A straightforward interpretation of such a language-specific weighting of dimensions would be that pitch and duration contribute differentially to the identification of the prominences on which the perception of contrastive rhythm is based. More intriguingly, the finding could re-introduce a hint of coordinative rhythm, specifically in the case of the Swiss-German speakers with their greater sensitivity to durational manipulations disrupting the interval between accents, suggesting that languages exhibit different balances of contrastive and coordinative rhythm. A way to test this would be to create stimuli like the ones described above, but with sufficient compensatory shortening or lengthening in the unaccented syllables following the manipulated accent that timing of the next accent is not perturbed relative to the original. However, before we commit to an interpretation entailing an overarching speech rhythm, we should remember that the language-specific results are also consistent with differential sensitivities to purely local events—segmental durations (German, but not French, has contrastive vowel length), and pitch dynamism as a cue to accent. Cumming's [50] language-specific weighting of duration and F0 in a combined PVI reduces the difference between Swiss-French and Swiss-German compared to a duration-only index, which supports the possibility that languages' superficial rhythmic differences depend on the weight they assign to different dimensions (cf. [53]; see also [44] for the suggestion that the relative strength in a given language of syllabic and foot-sized quasi-periodicities derived from the amplitude envelope may be crucially different between languages). It is just possible, therefore, that finding

the correct balance of prominence cues for each language will allow a universal underlying rhythmic character of speech to emerge from the surface and interlinguistic variation evident from the measurement of individual dimensions; but, as argued in §7, there are reasons not to expect language to submit to the restrictiveness of rhythmicality in its phonetic realization.

6. Rhythm versus linguistic structure

The apparent lack of isochrony in speech does not imply that speakers do not plan timing ahead. Speakers seem to exercise some sort of top-down control over segment or syllable duration and have a tendency to avoid having a too long or short linguistic unit [54]. In addition, the duration of pause within an utterance is affected by not only the length of its preceding phrase but also that of the following phrase in read speech [55,56].

Although it is difficult to tease apart the numerous factors directly affecting speech timing, studies on polysyllabic shortening show that the size of higher level units can affect the duration of lower level units, and this can still be understood as a temporal compensation process that may appear superficially as an adjustment towards isochrony between successive units (see [57–59] and also [8] for a review). For example, a central prosodic unit in Korean is the accentual phrase (AP)—rather infelicitously named, as we will see later that the language is a stranger to accentual prominence. Cues to the AP, which typically consists of three or four syllables [60], include adjustments to both pitch and duration (see e.g. [61]). In an experiment which bears on whether prosodic units adjust in a way which supports a tendency to any compensatory temporal adjustment (i.e. a tendency to lengthen syllables in smaller size AP) in addition to pitch, Jeon [61] used sequences of five or seven syllables which would have different meanings depending whether they were phrased (respectively) $2 + 3$ versus $3 + 2$, or $3 + 4$ versus $4 + 3$. The results regarding the pitch contour were straightforward: speakers produced a pitch leap between successive syllables across the AP boundary. In addition, the location of the AP boundary (i.e. the size of each accentual phrase) affected how much duration speakers assigned to each syllable within the utterance. When normalized syllable duration was compared between the two types of phrasing (Early Boundary [$2 + 3$ or $3 + 4$] versus Late Boundary [$3 + 2$ or $4 + 3$]), the general trend was that syllables in the smaller size AP tended to be lengthened compared with those in the AP with more syllables.

The presence of interacting timing-controllers is assumed in the coupled oscillator models [62,63]. In this approach, speech timing is modelled with multi-level oscillators of, for example, the syllable and the foot, although they can be extended to other linguistic units. Each oscillator has its own eigenfrequency; the coordinative rhythm is inherent in the timing-controllers. The lower level cycles are synchronized within the higher level cycle and the hierarchical cycles are repeated. The coupling strength between the cycles is parameterized and the setting of parameters in the model determines the temporal properties of the output signal. The models are based on theoretical assumptions that all languages have multi-level prosodic hierarchy but the contribution of each level to speech timing varies continuously across languages (cf. the discussion of the prominence gradient in §4). These models are successful in producing signals with complex durational variations as in natural speech with varying degrees of coordinative or contrastive rhythmicity. What is of importance is that the surface timing as successful outcome of the modelling process is not periodic (see also [64]), and the presence of the periodic temporal-controller is a hypothesis yet to be tested.

What is questionable is whether any compensatory timing process has any linguistically significant function, particularly when its occurrence is spread over several syllables—a 'diffuse' as opposed to 'localized' timing effect in the terms used by White [2]. He argues that the timing adjustments that are observed in speech can be accounted for not by a goal of making prosodic units at some level isochronous, that is, by mediation from a higher rhythmic plan, but purely by mechanisms to do with durational adjustments associated with marking the edges or heads (e.g. accented syllables) of prosodic domains. Once such functional durational adjustments (predominantly lengthening) are taken into account, he finds little evidence in the literature for compensatory shortening of the kind required to achieve isochrony in the face of variable amounts of phonetic material in domains such as the foot. Perception experiments on Korean [65] indeed revealed that localized lengthening on the potential phrase-final syllable serves as a more robust cue to the upcoming AP boundary than lengthening occurring over the phrase. That is, listeners are more efficient at exploiting the syntagmatic contrast of the pre-boundary syllable with what is before and after than they are at making use of less localized timing adjustments. Furthermore, in perception, as pointed out by Turk & Shattuck-Hufnagel [4], recovery of putative cyclical rhythms from the signal, which must involve discarding surface irregularity, must be done in a way 'which does not interfere with' the exploitation of these variations for grammatical purposes. In the next section, we consider the possibility that the search for speech rhythm does an injustice to the very nature of language.

7. Language and languages: rhythmic, arhythmic or antirhythmic?

We have seen that half a century of empirical research has conclusively demonstrated that even those languages seen as archetypes of syllable- or stress-timing are very far from exhibiting isochrony. We therefore have to retreat from any hope that languages are rhythmic in the everyday sense of clock-rhythmic. Can we then accept that the relevant notion of rhythm is *contrastive* rhythm, and that languages strive to achieve sequential alternation of prominences (albeit with less or more salient 'prominence gradients')?

If so, we must conclude that languages are doing a pretty poor job. First of all, what better way to smooth the path to contrastive rhythm than to develop strongly cued stressed syllables, as for instance in the Germanic languages. From that perspective, the question 'where's the stress in that word?' is sensible to ask, and easy to answer. Not so for speakers of many languages such as Tamil, Mongolian, Malay, West Greenlandic and Korean—or for the researchers who try to pin down what turns out to be a highly elusive property (e.g. [66–71]). The failure to develop this most useful foundation for contrastive rhythm suggests that such languages really are not trying very hard when it comes to rhythm. We

may just have to accept that some languages are content to remain relatively arhythmic—and yet, interestingly, they manage perfectly well to perform the communicative functions of language.

Second, and in case the languages with strongly cued stress are feeling smug at having mastered rhythm, once the spotlight falls on them it reveals they deserve no better than a B− for effort. The problem stems from that rather neglected aspect of language, syntagmatic contrast—one which has always been a fly in the ointment of attempts to impose on the analysis of speech the kind of cyclicity or rhythmicity of other activities such as walking or chewing [33]. Spoken English is akin to traffic flow, where we might expect, at any observation point and moment, an effectively random sequence of lorry-car-car-lorry-lorry-car-van-lorry-van-car-car-car rather than a rhythmically alternating lorry-car-lorry-car-lorry-car. This is despite the fact that a very straightforward way, in addition to developing strong stress, exists for a language to achieve the putative goal of contrastive rhythm, namely to incorporate the following into its phonology: a constraint on stressed syllables all to have the same structure, e.g. CVV (one consonant plus a long vowel); a constraint on unstressed syllables all to have the same structure as each other, e.g. CV (one consonant and a short vowel); and a constraint on the foot (the domain of a stressed syllable plus any following unstressed syllables) to be disyllabic. Sadly, even the languages which know how to recruit prominence to the goal of contrastive rhythm seem not to have had their eye on the ball when it comes to regularizing their metrical properties. So an utterance such as the following has highly irregular sequential syllable structures and foot structures:

he struggled to perceive the term 'nuclear'
hi | ˈstrʌ.gəld.tə.pə | ˈsiːv.ðə | ˈtɜːm | ˈnjuː.kli.ə
CV | ˈCCCV.CVCC.CV.CV | ˈCVVC.CV | ˈCVVC | ˈCCVV.CCV.V

In whatever way some might wish to tweak the metrical analysis (for instance by abandoning the Abercrombian foot [7] for structures which better reflect durational adjustments (cf. [72]), or use sleight-of-hand devices such as 'extrametricality' to dispose of supernumerary material), no amount of tampering will yield neat objective alternations of prominence in the signal. Words such as 'nuclear' (assuming three syllables, which is certainly possible) or 'polymer' [ˈpɒ.lɪ.mə] are particularly disobliging as they appear from their vowel quality to have (in non-rhotic English, at least) three syllables with decreasing prominence rather than any kind of alternation. Similarly, the Russian pronunciation of 'Gorbachëv' as [gərbɐˈtɕof] has a progressive rise in vowel prominence rather than alternation. Here, the vowel in the immediately pre-stress position, [ɐ], is the neutralization of /a/ and /o/, but in other unstressed positions the realization is further mid-centralized to [ə], which is also shorter than the pre-stress realization (according to [73]) in non-palatalizing environments, and no longer, at least, in palatalizing environments.

As all languages have had a long time to sort themselves out, we have to ask whether, in the case of a language like English, the blatant disregard for proper sequential alternation in favour of syntagmatic irregularity perhaps merits not only the term 'arhythmic', implying a degree of negligence on the part of the language, but even 'antirhythmic', redolent of wilful and rebellious disregard for decent metrical principles ('antirhythmic' was previously used by Pointon [74], but specifically of Spanish rather than in a general comment on the nature of languages). Or, maybe, those metrical principles of alternation have been overstated. Interestingly Morse Code, that analogy for English-type speech rhythm (usually attributed to Lloyd James [75]), which has often been assumed to be synonymous with stress-timing, is actually also antirhythmic in the sense of crucially depending on syntagmatic contrast. Letters are coded by one (• E, − T) two (e.g. •• I, −• N), three (e.g. ••− U, ••• S) or four (e.g. ••−• F, −•−− Y) dots or dashes (short or long beeps or flashes). Neither the letters, nor orthographically similar words they make up, have a fixed time slot. TEN [− • −•] and FUN [••−• ••− −•] take up different amounts of time, given that the principle of syntagmatic contrast in time is paramount, much as the syntagmatic contrasts in duration between the successive syllables in the example above support lexical distinctiveness.

8. Speech rhythm as metaphor

At this point—to the extent that the arguments for the arhythmicity or antirhythmicity of speech are persuasive—the question arises of how we ever came to apply the term 'rhythm' to speech and why its use is so widely accepted. Not only does speech uncontroversially lack strictly coordinative rhythm (it's not 'clock-rhythmic'), but it even shows a remarkably half-hearted approach to the lesser goal of contrastive rhythm.

The answer lies in our human ability for metaphorical extension. Think of the metaphor of a chessboard—which demonstrates perfect coordinative (here, spatial) and contrastive (black versus white) rhythm in two dimensions—as it might be applied to field patterns as viewed from the air. In most cases (excluding perhaps territories mapped out in recent times such as the American 'Mid-West') the fields will neither be perfectly regular in shape and size nor discretely of two different colours, but the metaphor is adequate to convey an impression of partial alternation of colour and apparent non-randomness of dimensions. As one moves through a continuum of landscapes to, at the other extreme, irregularly shaped patches of cultivation carved out of more rugged terrain, the metaphor will become less and less appropriate, to the point where it is unhelpful. The point of a metaphor is that we claim sameness between two elements which are self-evidently distinct, but by doing so draw attention to one or more properties of the 'source' of the metaphor which, by some stretch of the imagination, can be considered similar to properties in the 'target'. When we look out of the plane window in response to our companion drawing attention to 'that chessboard stretching out to the horizon', we do not expect to see a perfectly geometric pattern of black and white; but we can extract some properties of a chessboard (regular geometric shapes, alternation of black and white) which have an approximate parallel in properties of the landscape (quadrilateral field shapes, albeit less regular, and alternation of colours lighter and darker in a way analogical to the black and white of a chessboard). But, crucially, the applicability of the metaphor is not evidence for the action of an external template of regularity. The 'rhythm' of the fields arises from independent constraints—such as the optimal size and shape of a field, the lie of the land, ownership, and the need to rotate crops—and is merely 'emergent'.

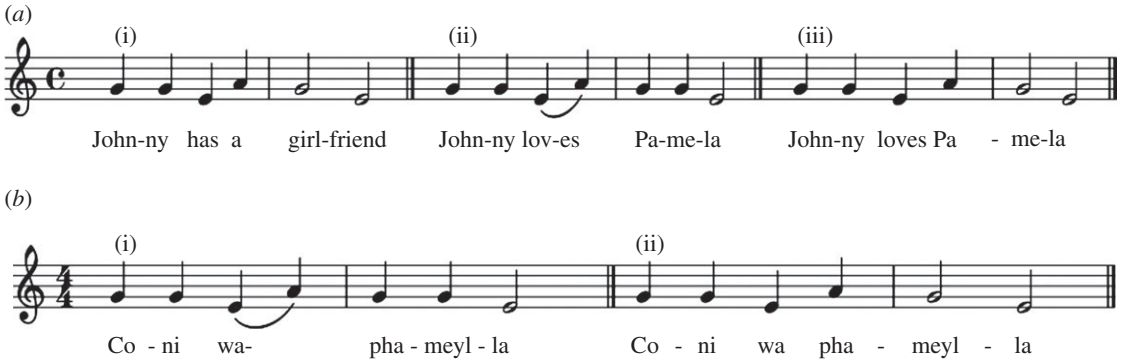

Figure 1. Well-formed and ill-formed tune–text associations. In (*a*), (i) and (ii) are well-formed, but (iii) is problematic for English speakers, whereas its Korean equivalent (*b*)(ii) is as natural as (*b*)(i).

Under this conception, likewise, speech is 'rhythmical' by metaphorical extension, and the characteristics which allow the metaphor to be applied emerge from independent aspects of the structure of language such as prominence cuing and domain-edge marking (cf. [2]). A language with salient prosodic prominences promotes analogy with the regularly spaced 'beats' of music or other periodic events, even though the occurrence of stressed and unstressed syllables is not constrained to be regular in time or to manifest strict alternation. A language with a shallower 'prominence gradient' may be harder to apply the rhythm metaphor to, but there will still be aspects of prominence and phrasing which can be drafted into service.

Arguing that speech is actually antirhythmic by virtue of such design features as syntagmatic contrast and length-based oppositions, and that it is therefore rhythmical only in a metaphorical sense, does not mean abandoning the quest to understand the relation of speech to rhythm; instead, it requires some re-framing of research questions about the rhythm of a particular language as questions about how the metaphor is applied, for instance, how it is that speech can be aligned in non-arbitrary ways to fundamentally rhythmic phenomena such as a metronome (as in speech cycling experiments, e.g. [76–78]) and of course, *par excellence*, music (we ignore here the fact that performance styles, expressivity, and the like may cause minor deviations from strict tempo). In speech cycling, speakers repeat a short utterance in time with repeating tones in experiments which are commonly periodic like metronome clicks, but the organization of the tones can vary as in a sequence of high-low-high-low-… or a waltz rhythm [76–78]. This paradigm forces speakers to align the utterance to the tones and the way speakers form a prominence-headed foot or a foot-like unit within the phase of clicks is analysed. Results show that, for example, English speakers have a strong preference for aligning stressed syllables to the tones [76], Korean speakers tend to align the accentual-phrase initial syllables [78], while speakers of Spanish or Italian find the task uncomfortable [77]. That is, in English, stressed syllables map onto beats, unstressed syllables can be phonetically reduced to force the alignment, but function words which are commonly unstressed (e.g. *for*) can be promoted as downbeats (i.e. the initial syllable in the foot). It is not entirely clear why Spanish or Italian speakers failed in performing the task despite the presence of lexical stress, but it may be due to the irreducibility of unstressed vowels (cf. [78]) or the presence of geminates or long vowels in the speech materials which are potential constraints on the formation of the foot in these languages. For Korean, Chung & Arvaniti [78] shows that the accentual-phrase initial syllable can map onto the beat, but Korean speakers would show more flexibility when facing such a forced-alignment task as shown in the musical example discussed below.

When language and music meet, then, cultures agree that there are right and wrong ways of associating them, but find language-specific solutions which may be less or more restrictive. To take a simple case which has been discussed in the intonational literature [79], the children's chant in figure 1*a* is only well formed in English if the beginning of both bars is associated with lexical stresses, as in (i) and (ii). Alternative (iii) is ill formed because it violates this requirement. If, however, the chant is translated into Korean, as in figure 1*b*, using the equivalent of 'Johnny' and 'Pamela', (*coni* ('Johnny') *wa* ('and') *phameylla* ('Pamela'), /tɕoniwa pʰamɛlla/), both associations (i) and (ii) (equivalent to (ii) and (iii) in figure 1*a*) are well formed and are equally comfortable for a Korean speaker. The languages clearly show a difference of strategy with respect to musical association, and so the research question is what the distinct properties of different languages are which can be recruited to fulfil the application of the rhythm metaphor.

Surely, though, such alignment is a matter of fitting one rhythm to another? Maybe Korean utterances just contain two competing rhythms, either of which can be aligned with the music. But it need not be the case that when we fit one thing to a second which is rhythmic the first must also be rhythmic. Imagine an archaeologist finding an unbound ancient manuscript 10 inches (in.) thick. It is in a completely unknown language and writing system. To protect and preserve it for transport, we have to fit it into a special small airtight filing cabinet with 10 drawers each 2 in. high and a smaller gap between each drawer—a 'coordinative-rhythmic' item of archaeological equipment if ever there was. Now, we might adopt a simple strategy of dividing the manuscript into ten 1-in. sections, one for each drawer, with plenty of space to spare. Alternatively, we might notice that every so often within the manuscript a page might have fewer, much larger inscriptions, and maybe some illuminations—rather like a title page. We hypothesize that these may have significance; that is, the manuscript apparently reflects some kind of semantic, informational or ceremonial *structure*. As, by the serendipity of thought-experiments, there are 10 such special pages, we instead divide the manuscript into 10 'chapters', one per drawer, even though the largest only just fits into the 2-in. drawer and the smallest chapters are only a few pages. Was the manuscript rhythmic? Have we matched two rhythms? We would contend that the manuscript had *structure*, but not *rhythm*—because, as to a considerable extent with modern

book chapters, there is no requirement for regularity in the elements of structure, nor any necessary pattern in their sequencing such as long-short-short. Whatever the reasons for the structural divisions, they are unlikely to have been motivated by a goal of rhythm in the manuscript.

Similarly, we espouse the view here that there is no goal of rhythmicality in speech. Rather, timing is the servant of linguistic structure, including lexical differentiation (in which for instance length, stress and syllabic complexity variously play roles in different languages), prominence for informational purposes and prosodic edge-marking as an aid to parsing utterances. If speech is not teleologically rhythmic, far from making tune–text association less interesting it makes it all the more impressive as it involves not a mapping between two entities in the same domain (that of rhythm), but a metaphorical analogy between the rhythm of the target (the music, the chant, the metronome in speech-cycling experiment) and some non-rhythmically motivated aspect of the structure of the speech. In the case of a language such as English with a steep prominence gradient, the metaphorical relation between points of rhythmical strength in the target and linguistically determined prominences may lead to a straightforward, and potentially unique, mapping. In a language such as Korean lacking such clear prominences, the metaphor fits less well, and the alignment solution is less determinate. We would also predict that in the case of entrainment experiments, where speakers are forced into applying the rhythm metaphor and finding analogies to an explicit beat and implicit subdivisions, there will be greater diversity of solution for languages without a steep prominence gradient.

9. Conclusion

We have discussed a subset of matters relating to the definition, and indeed existence, of rhythm in speech. This is of course a potentially circular discussion, as how tightly or loosely one defines rhythm will affect whether or not spoken language satisfies that definition. But in respect of two plausible conceptualizations of rhythm, we can agree that everyday speech does not appear to have coordinative rhythm with its implied synchronization of one or more unit to an external clock, and that the notion of contrastive rhythm, based on alternation between units of different prominence, is the one that has the better potential to model the facts of speech. For a language such as English, where alternation between stronger and weaker elements is most likely to be found, by virtue of there being very clear cues to stressed and unstressed syllables, we have however noted that sustaining the concept of 'alternation' (or cyclicity) requires turning a blind eye to the actual phonetic properties of the sequences of syllables and feet that occur. English allows a rich variety of syntagmatic contrast and achieves a poor approximation to alternation (as critics of the PVI are happy to point out), even though relatively simple mechanisms could have been applied in its historical development that would have culminated in rhythmic rectitude. Our view, then, is that the simulacrum of rhythm in speech is accidental. White [2] shows how durational adjustments for functions such as marking domain edges and heads may in some cases simulate goal-oriented compression to achieve rhythm; and the apparent alternation of prominences of strong-stress languages is merely emergent from the way the lexicon is partitioned (in part by stress and in part by flexibility in the length of words), and from happenstance when words with their inherent stresses are concatenated.

Even if we were to be permissive in our acceptance of what constitutes 'alternation' for those languages such as English, German, Russian and so on which manifest strongly cued stress, we still have to contend with a number of languages (such as Tamil, Mongolian, Malay and Korean) where the lexicon apparently fails to specify a given point in a word as the recipient of culminative prosodic prominence, so that in words or continuous utterances it is remarkably hard to identify any syllable-by-syllable alternation of prominence. Information structure may determine greater prominence on certain elements at higher levels in an utterance, but the distribution of these prominences will be primarily determined by semantic, lexical and pragmatic factors rather than a goal of achieving rhythm.

Perhaps, then, speech has neither coordinative nor contrastive rhythm, and so is constrained neither to synchronize with an external clock pulse, nor to achieve a sequential alternation of prominence. In this, we mirror the view expressed in Turk & Shattuck-Hufnagel [4] that alternative models need to be considered 'in which periodicity plays no role in normal conversational speech'. If this turns out to be the case it does not, however, prevent speakers of a language applying the *metaphor* of rhythm, just as the chessboard metaphor can be applied to field patterns seen from the air—particularly when obliged to do so when matching speech to external rhythms. The fit between the metaphor and reality will never be perfect and its application will vary from more appropriate to less appropriate across a range of languages respectively (just as the chessboard metaphor fits some landscapes better than others). The metaphor mediates the association of text with music—more strictly in those languages which fit the metaphor more comfortably, and more permissively (or ambiguously) in languages where prominences are less salient.

Our purpose in this paper has been to step aside for a moment from the familiar endeavour of trying to elucidate the nature of rhythm in speech and to adopt a different viewpoint, one which questions the presupposition that there is a rhythm to elucidate. Doing so might be compared to, in the realm of religion, departing from a strict theistic stance and adopting an agnostic one. Rather than taking a phenomenon (God or speech rhythm) as axiomatic, and trying to refine our understanding of its nature, this alternative stance frees us to ask new questions and reformulate existing ones. This agnostic stance certainly does not preclude us suggesting refined tests which might better reveal a rhythm immanent in speech, as with our discussion (§5) of the language-specific weighting of different dimensions in a complex rhythm metric. It obliges us, however, also to focus on the consequences of the alternative hypothesis, i.e. that speech is not inherently rhythmical. For instance, we need to think about the definition of arhythmicality as well as rhythm; for unless there are *potential* sequences of elements varying in duration and salience which can unambiguously be described as lacking rhythm within the relevant framework, then the concept of rhythm becomes vacuous. Questions about the association of speech to incontestibly rhythmical activities will be framed differently, in terms not of matching two rhythms but of finding the best analogues in the linguistic signal to allow the metaphorical interpretation in speech of the rhythm in question. As those analogues will depend

in part on the phonology (including prosody) of the given language, and as there is wide variation in phonology across languages, the metaphor will select different properties language specifically, and indeed will be easier to apply, or to apply unambiguously, in some languages than others. To the extent that this is true, it suggests spoken language is more the handmaiden of language as an abstract cognitive system, and is less the slave of cyclical physical activities (comparable to walking). If, on the other hand, we reject the alternative perspective, and return to the notion that speech is inherently rhythmical, the onus is to explain why languages do not reflect the fact more transparently. As we noted, there are easy remedies by way of historical change which could have achieved this by now, but they have not been implemented. What is clear is that progress in understanding the rhythm of speech—or the lack of it—will be best served by including the widest and most diverse range of languages in research within all the relevant paradigms.

Acknowledgments. The authors thank two anonymous reviewers and the editor, Tamara Rathcke, for their thought-provoking and very helpful comments.

References

1. White L, Mattys SL, Wiget L. 2012 Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *J. Mem. Lang.* **66**, 665–679. (doi:10.1016/j.jml.2011.12.010)
2. White L. 2014 Communicative function and prosodic form in speech timing. *Speech Commun.* **63-64**, 38–54. (doi:10.1016/j.specom.2014.04.003)
3. Couper-Kuhlen E. 1986 *An introduction to English prosody.* London, UK: Arnold.
4. Turk A, Shattuck-Hufnagel S. 2013 What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Lab. Phonol.* **4**, 93–118. (doi:10.1515/lp-2013-0005)
5. Patel AD. 2008 *Music, language, and the brain*. Oxford, UK: Oxford University Press.
6. Pike KL. 1945 *The intonation of American English.* Ann Arbor, MI: University of Michigan Press.
7. Abercrombie D. 1967 *Elements of general phonetics.* Edinburgh, UK: Edinburgh University Press.
8. Fletcher J. 2010 The prosody of speech: timing and rhythm. In *The handbook of phonetic sciences* (eds WJ Hardcastle, J Laver, FE Gibbon), pp. 523–602, 2nd edn. Chichester, UK: Wiley-Blackwell.
9. Kohler K. 2009 Rhythm in speech and language. *Phonetica* **66**, 29–45. (doi:10.1159/000208929)
10. Lehiste I. 1977 Isochrony reconsidered. *J. Phon.* **5**, 253–263.
11. Lehiste I. 1979 The perception of duration within sequences of four intervals. *J. Phon.* **7**, 313–316.
12. Dauer RM. 1983 Stress-timing and syllable-timing reanalyzed. *J. Phon.* **11**, 51–62.
13. Eriksson A. 1991 Aspects of Swedish speech rhythm. *Gothenburg Monogr. Linguist.* **9**.
14. Arvaniti A. 2009 Rhythm, timing and the timing of rhythm. *Phonetica* **66**, 46–63. (doi:10.1159/000208930)
15. Ramus F, Nespor M, Mehler J. 1999 Correlates of linguistic rhythm in the speech signal. *Cognition* **73**, 265–292. (doi:10.1016/S0010-0277(99)00058-X)
16. Ramus F, Dupoux E, Mehler J. 2003 The psychological reality of rhythm classes: perceptual studies. In *Proc. 15th Int. Congr. Phonetic Sciences, 3–9 August 2003, Barcelona, Spain*, pp. 337–342. Barcelona, Spain: Universitat Autònoma de Barcelona.
17. Dellwo V, Wagner P. 2003 Relations between language rhythm and speech rate. In *Proc. 15th Int. Congr. Phonetic Sciences, 3–9 August 2003, Barcelona, Spain*, pp. 471–474. Barcelona, Spain: Universitat Autònoma de Barcelona.
18. Dellwo V. 2006 Rhythm and speech rate: a variation coefficient for deltaC. In *Language and language processing: proceedings of the 38th Linguistic Colloquium* (eds P Karnowski, I Szigeti), pp. 231–241. Frankfurt, Germany: Peter Lang.
19. Tongue RK. 1974 *The English of Singapore and Malaysia*. Singapore: Eastern Universities.
20. Platt JT, Weber H. 1980 *English in Singapore and Malaysia: status, features and functions*. Kuala Lumpur, Malaysia: Oxford University Press.
21. Low EL. 1998 Prosodic prominence in Singapore English. PhD dissertation, University of Cambridge.
22. Low EL, Grabe E, Nolan F. 2000 Quantitative characterizations of speech rhythm: 'syllable-timing' in Singapore English. *Lang. Speech* **43**, 377–401. (doi:10.1177/00238309000430040301)
23. Grabe E, Low EL. 2002 Durational variability in speech and the rhythm class hypothesis. In *Papers in laboratory phonology VII* (eds C Gussenhoven, N Warner), pp. 515–546. The Hague, The Netherlands: Mouton de Gruyter.
24. Dankovičová J, Dellwo V. 2007 Czech speech rhythm and the rhythm class hypothesis. In *Proc. 16th Int. Congr. Phonetic Sciences, 6–10 August 2007, Saarbrücken, Germany*, pp. 471–474. Saarbrücken, Germany: Universität des Saarlandes.
25. Mok P. 2009 On the syllable-timing of Cantonese and Beijing Mandarin. *Chin. J. Phon.* **2**, 148–154.
26. Nolan F, Asu EL. 2009 The pairwise variability index and coexisting rhythms in language. *Phonetica* **66**, 64–77. (doi:10.1159/000208931)
27. Arvaniti A. 2012 The usefulness of metrics in the quantification of speech rhythm. *J. Phon.* **40**, 351–373. (doi:10.1016/j.wocn.2012.02.003)
28. O'Rourke E. 2008 Speech rhythm variation in dialects of Spanish: applying the pairwise variability index and variation coefficients to Peruvian Spanish. In *Proc. Speech Prosody Conf., 6–9 May 2008, Campinas, Brazil*, pp. 431–434. Campinas, Brazil: State University of Campinas.
29. Carter PM. 2005 Quantifying rhythmic differences between Spanish, English, and Hispanic English. In *Theoretical and experimental approaches to romance linguistics: selected papers from the 34th Linguistic Symposium on Romance Languages* (eds RS Gess, EJ Rubin), pp. 63–75. Amsterdam, The Netherlands: John Benjamins.
30. Deterding D. 2001 The measurement of rhythm: (a comparison of Singapore and British English. *J. Phon.* **29**, 217–230. (doi:10.1006/jpho.2001.0138)
31. Gibbon D. 2003 Computational modelling of rhythm as alternation, iteration and hierarchy. In *Proc. 15th Int. Congr. Phonetic Sciences, 3–9 August 2003, Barcelona, Spain*, pp. 2489–2492. Barcelona, Spain: Universitat Autònoma de Barcelona.
32. Knight R-A. 2011 Assessing the temporal reliability of rhythm metrics. *J. Int. Phon. Assoc.* **41**, 271–281. (doi:10.1017/S0025100311000326)
33. Nolan FJ. 1982 The role of action theory in the description of speech production. *Linguistics* **20**, 287–308. (doi:10.1515/ling.1982.20.3-4.287)
34. Lin H, Wang Q. 2007 Mandarin rhythm: an acoustic study. *J. Chin. Linguist. Comput.* **17**, 127–140.
35. Nazzi T, Ramus F. 2003 Perception and acquisition of linguistic rhythm by infants. *Speech Commun.* **41**, 233–243. (doi:10.1016/S0167-6393(02)00106-1)
36. White L, Mattys S, Series L, Gage S. 2007 Rhythm metrics predict rhythmic discrimination. In *Proc. Int. Congr. Phonetic Sciences XVI, 6–10 August 2007, Saarbrücken, Germany*, pp. 1009–1012. Saarbrücken, Germany: Universität des Saarlandes.
37. Ramus F, Mehler J. 1999 Language identification with suprasegmental cues: a study based on speech resynthesis. *J. Acoust. Soc. Am.* **105**, 512–521. (doi:10.1121/1.424522)
38. Nazzi T, Bertoncini J, Mehler J. 1998 Language discrimination by newborns: towards an understanding of the role of rhythm. *J. Exp. Psychol.* **24**, 756–766. (doi:10.1037/0096-1523.24.3.756)
39. Ramus F. 2002 Acoustic correlates of linguistic rhythm: perspectives. In *Proc. Speech Prosody Conf., 11–13 April 2002, Aix-en-Provence, France*, pp. 115–120. Aix-en-Provence, France: Laboratoire Parole et Langage, Université de Provence.
40. Klatt DH. 1976 Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *J. Acoust. Soc. Am.* **59**, 1208–1220. (doi:10.1121/1.380986)

41. Wenk BJ, Wioland F. 1982 Is French really syllable-timed? *J. Phon.* **10**, 193–216.
42. Cumming RE. 2011 The language-specific interdependence of tonal and durational cues in perceived rhythmicality. *Phonetica* **68**, 1–25. (doi:10.1159/000327223)
43. Christophe A, Morton J. 1998 Is Dutch native English? Linguistic analysis by 2-month-olds. *Dev. Sci.* **1**, 215–219. (doi:10.1111/1467-7687.00033)
44. Tilsen S, Arvaniti A. 2013 Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *J. Acoust. Soc. Am.* **134**, 628–639. (doi:10.1121/1.4807565)
45. Barry WJ, Andreeva B, Russo M, Dimitrova S, Kostadinova T. 2003 Do rhythm measures tell us anything about language type? In *Proc. 15th Int. Congr. Phonetic Sciences*, *3–9 August 2003, Barcelona, Spain*, pp. 2693–2696. Barcelona, Spain: Universitat Autònoma de Barcelona.
46. Loukina A, Kochanski G, Rosner B, Keane E, Shih C. 2011 Rhythm measures and dimensions of durational variation in speech. *J. Acoust. Soc. Am.* **129**, 3258–3270. (doi:10.1121/1.3559709)
47. White L, Mattys SL. 2007 Calibrating rhythm: first language and second language studies. *J. Phon.* **35**, 501–522. (doi:10.1016/j.wocn.2007.02.003)
48. White L, Mattys SL. 2007 Rhythmic typology and variation in first and second languages. In *Segmental and prosodic issues in Romance phonology* (eds P Prieto, J Mascaro, M-J Sole), pp. 237–257. Amsterdam, The Netherlands: John Benjamins.
49. Ferragne E. 2008 Étude phonétique des dialectes modernes de l'anglais des Îles Britanniques: vers l'identification automatique du dialecte. PhD dissertation, University of Lyon.
50. Cumming RE. 2011 Perceptually informed quantification of speech rhythm in pairwise variability indices. *Phonetica* **68**, 256–277. (doi:10.1159/000335416)
51. Koreman J, Andreeva B, Barry W. 2008 Accentuation cues in French and German. In *Proc. Speech Prosody Conf.*, *6–9 May 2008, Campinas, Brazil*, pp. 613–616. Campinas, Brazil: State University of Campinas.
52. Kochanski G, Grabe E, Coleman J, Rosner B. 2005 Loudness predicts prominence: fundamental frequency lends little. *J. Acoust. Soc. Am.* **118**, 1038–1054. (doi:10.1121/1.1923349)
53. Barry W, Andreeva B, Koreman J. 2009 Do rhythm measures reflect perceived rhythm? *Phonetica* **66**, 78–94. (doi:10.1159/000208932)
54. Jun S-A. 2005 Prosodic typology. In *Prosodic typology* (ed. S-A Jun), pp. 430–458. Oxford, UK: Oxford University Press.
55. Krivokapić J. 2007 Prosodic planning: effects of phrasal length and complexity on pause duration. *J. Phon.* **35**, 162–179. (doi:10.1016/j.wocn.2006.04.001)
56. Zvonik E, Cummins F. 2003 The effect of surrounding phrase lengths on pause duration. In *Proc. of Eurospeech Conf.*, *1–4 September 2003, Geneva, Switzerland*, pp. 777–780. Geneva, Switzerland: ISCA.
57. Turk AE, Shattuck-Hufnagel S. 2000 Word-boundary-related duration patterns in English. *J. Phon.* **28**, 397–440. (doi:10.1006/jpho.2000.0123)
58. White L, Turk A. 2010 English words on the Procrustean bed: polysyllabic shortening reconsidered. *J. Phon.* **38**, 459–471. (doi:10.1016/j.wocn.2010.05.002)
59. Lehiste I. 1980 Interaction between test word duration and length of utterance. In *The melody of language* (eds LR Waugh, CH Schooneveld), pp. 169–176. Baltimore, MD: University Park Press.
60. Jun S-A. 1998 The accentual phrase in the Korean prosodic hierarchy. *Phonology* **15**, 189–226. (doi:10.1017/s0952675798003571)
61. Jeon H-S. 2011 Prosodic phrasing in Seoul Korean: the role of pitch and timing cues. PhD dissertation, University of Cambridge.
62. O'Dell M, Nieminen T. 2001 Speech rhythms as cyclical activity. In *Papers from the 21st Meet. of Finnish Phoneticians* (eds S Ojala, J Tuomainen), pp. 159–168. Turku, Finland: Department of Finnish and General Linguistics of the University of Turku.
63. Barbosa PA. 2007 From syntax to acoustic duration: a dynamical model of speech rhythm production. *Speech Commun.* **49**, 725–742. (doi:10.1016/j.specom.2007.04.013)
64. Windmann A, Šimko J, Wrede B, Wagner P. 2012 Optimization-based model of speech timing and rhythm. In *Proc. Laboratory Phonology 13, 27–29 July 2012, Stuttgart, Germany*, pp. 175–176. Stuttgart, Germany: Association for Laboratory Phonology.
65. Jeon H-S, Nolan F. 2013 The role of pitch and timing cues in the perception of phrasal grouping in Seoul Korean. *J. Acoust. Soc. Am.* **133**, 3039–3049. (doi:10.1121/1.4798663)
66. Keane E. 2006 Prominence in Tamil. *J. Int. Phon. Assoc.* **36**, 1–20. (doi:10.1017/S0025100306002337)
67. Karlsson A. 2005 *Rhythm and intonation in Halh Mongolian*. Travaux de l'institute de linguistique de Lund 46. Lund, Sweden: Lund University.
68. Maskikit R, Gussenhoven C. 2013 No stress, no pitch accent, no prosodic focus: the case of Moluccan Malay. In *Paper presented at PaPI* (*Phonetics and Phonology in Iberia*) *2013*, 25–26 June, Lisbon, Portugal. Lisbon, Portugal: University of Lisbon.
69. Jacobsen B. 2000 The question of 'stress' in West Greenlandic: an acoustic investigation of rhythmicization, intonation, and syllable weight. *Phonetica* **57**, 40–67. (doi:10.1159/000028458)
70. Lim B-J. 2001 The production and perception of word-level prosody in Korean. *IULC Working Papers in Linguistics* **1**, 1–14.
71. Kim J-M, Flynn S, Oh M. 2007 Non-native speech rhythm: a large-scale study of English pronunciation by Korean learners. *Stud. Phon. Phonol. Morphol.* **13**, 219–250.
72. Shattuck-Hufnagel S, Turk A. 2011 Durational evidence for word-based versus prominence-based constituent structure in limerick speech. In *Proc. 17th Int. Congr. Phonetic Sciences*, *17–21 August 2011, Hong Kong*, pp. 1806–1809. Hong Kong, SAR China: International Phonetic Association.
73. Padgett J, Tabain M. 2005 Adaptive dispersion theory and phonological vowel reduction in Russian. *Phonetica* **62**, 14–54. (doi:10.1159/000087223)
74. Pointon GE. 1980 Is Spanish really syllable-timed? *J. Phon.* **8**, 293–304.
75. Lloyd James A. 1940 *Speech signals in telephony*. London, UK: Pitman & Sons.
76. Cummins F, Port R. 1998 Rhythmic constraints on stress timing in English. *J. Phon.* **26**, 145–171. (doi:10.1006/jpho.1998.0070)
77. Cummins F. 2002 Speech rhythm and rhythmic taxonomy. In *Proc. of Speech Prosody Conf.*, *11–13 April 2002, Aix-en-Provence, France*, pp. 121–126. Aix-en-Provence, France: Laboratoire Parole et Langage, Université de Provence.
78. Chung Y, Arvaniti A. 2013 Speech rhythm in Korean: experiments in speech cycling. *Proc. Meet. Acoust.* **19**, 060216. (doi:10.1121/1.4801062)
79. Ladd DR. 2008 *Intonational phonology*. Cambridge, UK: Cambridge University Press.